

# SenseGen: A Deep Learning Architecture for Synthetic Sensor Data Generation



Moustafa Alzantot (UCLA), Supriyo Chakraborty (IBM Research), Mani Srivastava (UCLA)

Large number of applications rely on personal sensory data. However, sharing personal sensor data may reveal a lot of sensitive information. To circumvent this issue, synthetic data can be used as an alternative to real data. Generated data should preserve only the required statistics of the real data to achieve the service offered by the data recipient, and should be hard to distinguish from real sensory data.

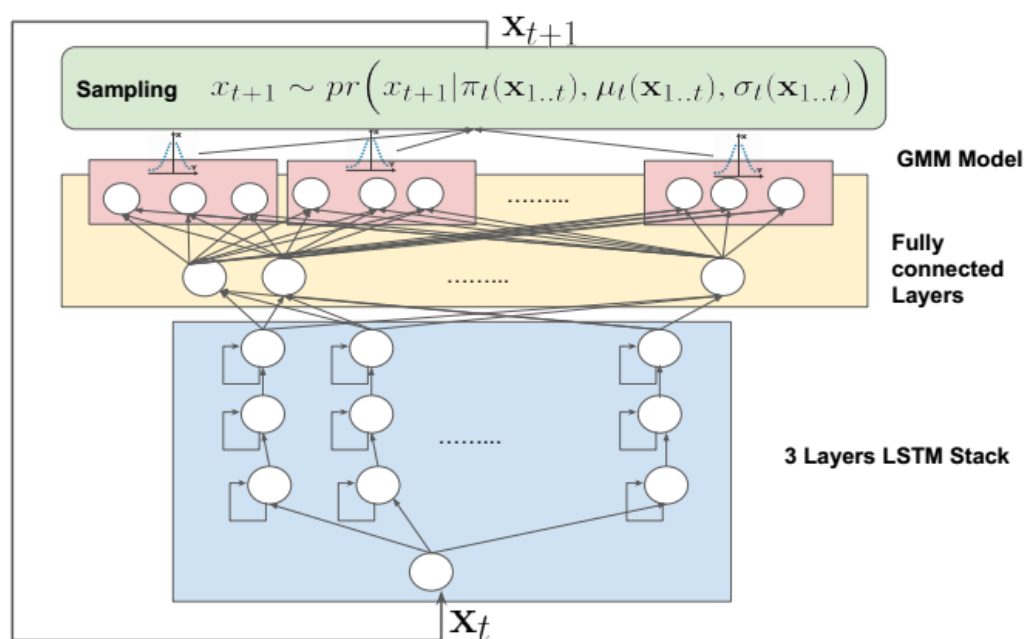
As a solution, we present, *SenseGen* a deep learning based architecture to synthesize sensory data.

## Model Architecture

SenseGen is composed of two models:

- **Generator model:** produces synthetic time series data from random noise inputs.
- **Discriminator model:** assesses the similarity between the generator output and the real sensor data.

Generator model is composed of a stack of long short term memory (LSTM) recurrent neural network layers, and mixture density network (MDN).



## Training Algorithm

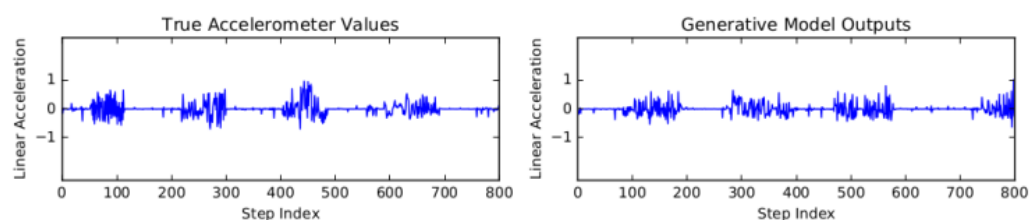
### Algorithm 1 Training algorithm

- 1: **for**  $t = 1, 2, \dots, T$  **do**
- 2: Sample  $\mathcal{X}_{true}$  minibatch from true data
- 3: Sample  $\mathcal{X}_{gen}$  minibatch from the generative model  $G$
- 4: Train the discriminative model  $\mathcal{D}$  on the training set  $(\mathcal{X}_{true}, \mathcal{X}_{gen})$  for 200 epochs
- 5: Sample another  $\mathcal{X}_{true}$  minibatch from true data
- 6: Train the generative model  $G$  on the training set  $(\mathcal{X}_{true})$  for 100 epochs
- 7: **end for**

## Results:

### Synthetic accelerometer data

Trained on accelerometer data from the Human Activity Recognition (HAR) database.



### Synthetic ECG data

